UCRL-ID-150775

Separation of Climate Signals

I. K. Fodor and C. Kamath

November 13, 2002

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

This work was performed under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

Separation of climate signals

Imola K. Fodor and Chandrika Kamath
SciDAC Scientific Data Management Center
Data Mining and Access Pattern Discovery Project

November 13, 2002

1 Introduction

Understanding changes in global climate is a challenging scientific problem. Simulated and observed data include signals from many sources, and untangling their respective effects is difficult. In order to make meaningful comparisons between different models, and to understand human effects on global climate, we need to isolate the effects of different sources.

Recent eruptions of the El Chichón and Mt. Pinatubo volcanoes coincided with large El Niño and Southern Oscillation (ENSO) events, which complicates the separation of their contributions on global temperatures. Current approaches for separating volcano and ENSO signals in global mean data involve parametric models and iterative techniques [3]. We investigate alternative methods based on principal component analysis (PCA) [2] and independent component analysis (ICA) [1]. Our goal is to determine if such techniques can automatically identify the signals corresponding to the different sources, without relying on parametric models. Fig. 1 summarizes our approach.

2 Results with synthetic data

ICA has been applied successfully in various source separation problems, such as removing artifacts from EEG/MEG data [3]. Fig. 2 illustrates the method with simple artificial data. In this case, ICA correctly estimates the shapes of two underlying sources S_1 and S_2 from the two mixed signals X_1 and X_2. As illustrated in Fig. 3, with proper post-processing, the amplitudes of the two sources can also be estimated accurately. Under the assumption of linear mixtures of independent signals, ICA performs well.

3 Results with climate data

We obtained monthly mean temperature re-analysis data from the National Centers for Environmental Prediction (http://www.ncep.noaa.gov/). The data is on

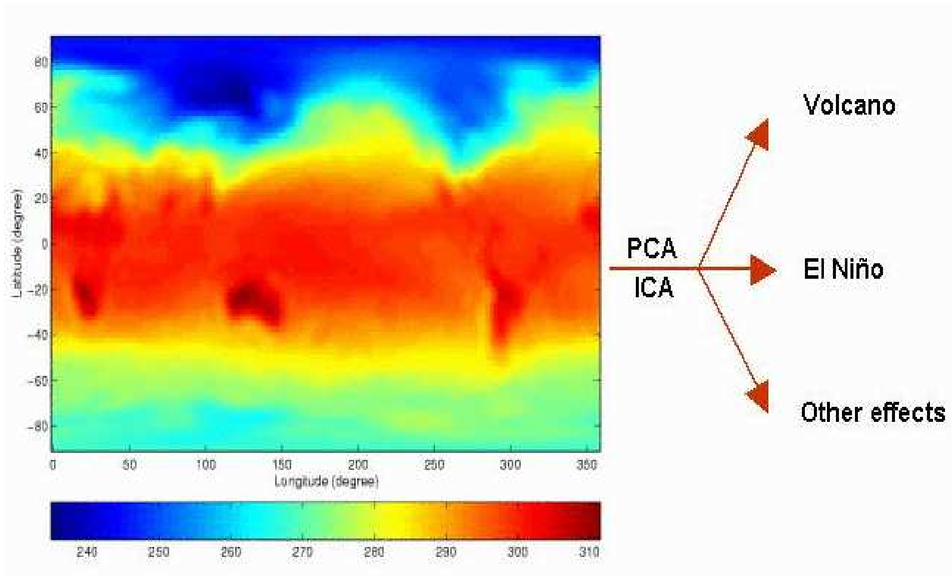

Figure 1: January 1979 raw temperatures (Kelvin) on the 144x73 latitude by longitude grid at 1000hPa pressure level.

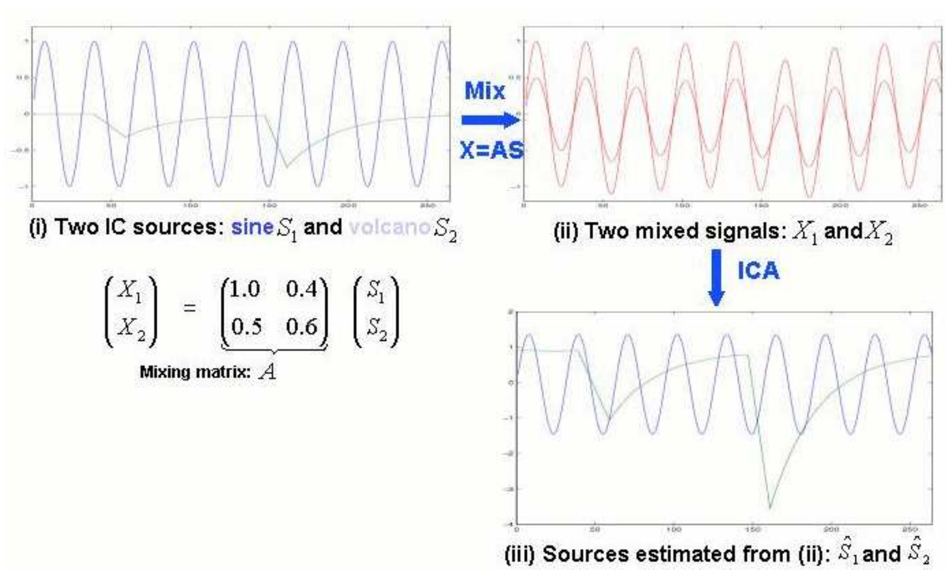

Figure 2: ICA applied to synthetic data.

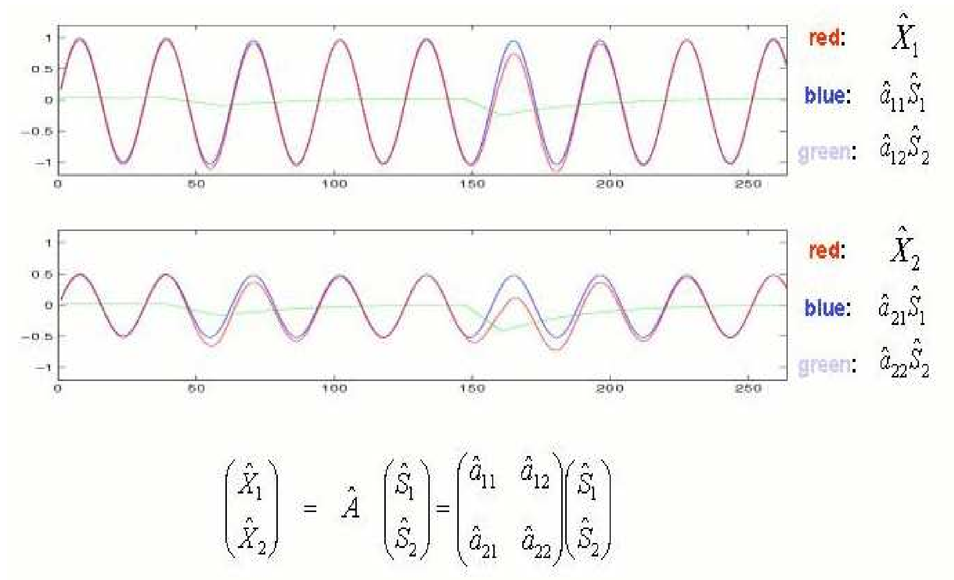

Figure 3: The mixed synthetic signals in terms of the estimated independent components.

a 144 x 73 longitude-by-latitude grid, on 17 vertical pressure levels (1,0000hPa close to the surface to 10hPa at the highest elevation), spanning 264 months (22 years, from 1979 to 2000). Since we expected ENSO and volcano signals to have strong latitudinal dependence, we performed our analyses on zonally averaged data. At a given month and level, we first calculated the 73 zonal means (over the 144 longitudes) at the 73 latitudes. Next, following standard practices in the atmospheric sciences [2], we removed the seasonal variation and centered the data as follows. For each month, we replaced the values at each of the 73 x 17 latitude-by-level grid points by subtracting their corresponding monthly means over the entire 22-year period. We therefore perform PCA and ICA on a 256 x 1241 time-by-space dimension anomaly dataset Z, weighted properly to account for the unequal spatial grid sizes.

Fig. 4 displays the first six basis functions obtained with PCA. The x-axis has 17 pressure levels, the y-axis has 73 latitude values.

Using PCA, we found that the first k=22 PCs explain more than 90% of the variation. In order to reduce the dimension of the ICA problem, we applied ICA to the k=22 PC basis functions. Since the PCs are simply linear combinations of the original data, we can easily obtain the IC coefficients corresponding to the anomaly data by linear transformations of the IC coefficients obtained for the PCs. Fig. 5 displays the resulting IC basis functions. Unlike the PC functions in Fig. 4, the IC functions in Fig. 5 are localized spatially. They do provide an independent basis representation for the data, but they are not interpretable in

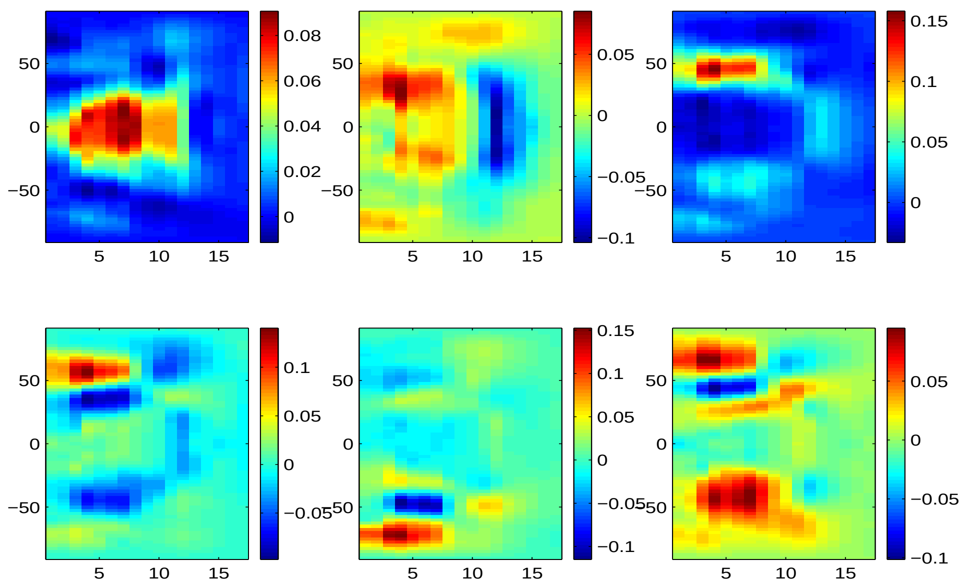

Figure 4: First 6 basis functions, obtained by applying PCA to the zonal anomaly data.

terms of atmospheric processes. Scientific interpretability is a requirement in our case, so next we explored more appropriate ICA models.

Fig. 6 displays the results of applying ICA to the anomaly data projected onto the first k=22 PC basis functions. In contrast to the previous ICA method in Fig. 5, the results of this version of ICA appear to be scientifically interpretable.

The next task is to determine which representation is better suited to separate the signals corresponding to the various sources. To answer that problem, we project the data onto the interpretable bases, and thus obtain the corresponding time series coefficients. Fig. 7 shows the projection of the anomalies onto the first 6 PC basis functions in Fig. 4. The main task is to identify which atmospheric processes the different time series represent.

4 Summary

We are currently investigating efficient methods to automatically separate signals in climate data. In addition to the widely used PCA in the climate community, we are also experimenting with the more novel ICA. Our initial results combining ICA with PCA are promising.

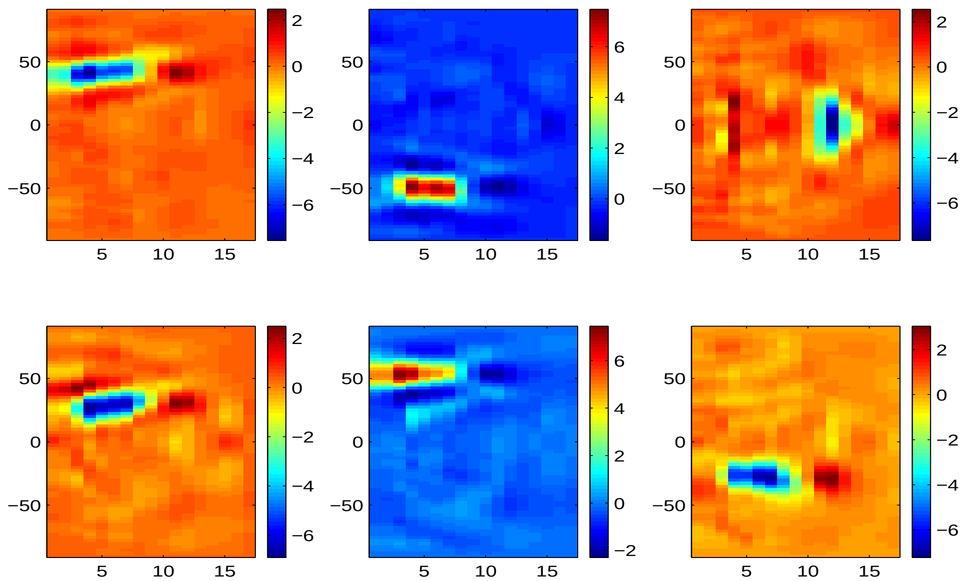

Figure 5: First 6 basis functions, obtained by applying ICA to the first k=22 PC basis functions.

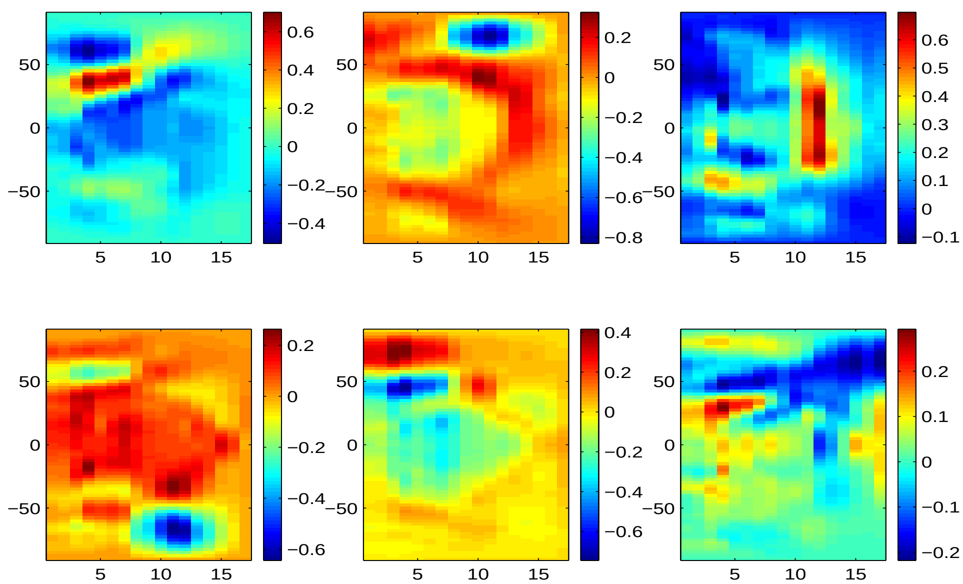

Figure 6: First 6 basis functions, obtained by applying ICA to the anomaly data projected onto the first k=22 PC basis functions (i.e. ICA applied to the PC time series coefficients).

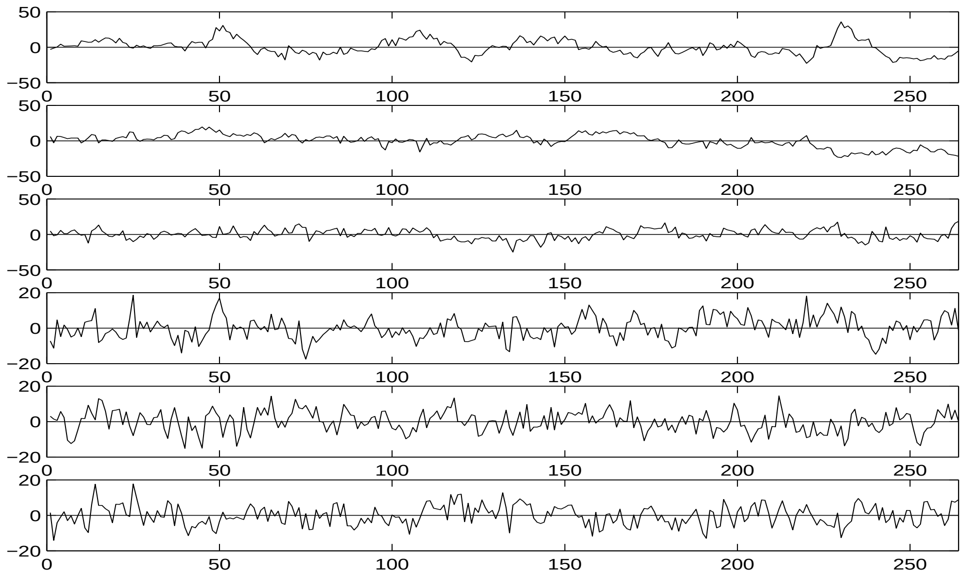

Figure 7: First 6 PC time series coefficients.

Acknowledgments

We thank Dr. Benjamin D. Santer from the Program for Climate Model Diagnostics and Intercomparison at Lawrence Livermore National Laboratory for introducing us to the problem, for providing the climate data, and for his numerous insights and suggestions.

UCRL-ID-150775. This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

Contact info

For more technical information, contact Dr. Chandrika Kamath at the Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-560 Livermore CA 94551, (925) 423-3768, kamath2@llnl.gov, http://www.llnl.gov/casc/sapphire/.

References

[1] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis.* Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control. Wiley, 2001.

[2] R.W. Preisendorfer. *Principal Component Analysis in Meteorology and Oceanography.* Number 17 in Developments in Atmospheric Science. Elsevier Science, Amsterdam, 1988.

[3] B.D. Santer et al. Accounting for the effects of volcanoes and ENSO in comparisons of modeled and observed temperature trends. *Journal of Geophysical Research*, 106(D22):28,033–28,059, November 27 2001.